November 1998

Assessment in Literature-Based Reading Programs: Have We Kept Our Promises?

Tanja Bisesi
Michigan State University

Devon Brenner
Michigan State University

Mary McVee
Michigan State University

P. David Pearson
Michigan State University

Loukia K. Sarroub
University of Nebraska-Lincoln, lsarroub@unl.edu

Published in:

Literature-Based Instruction: Reshaping the Curriculum
Edited by Taffy E. Raphael and Kathryn H. Au
Christopher-Gordon Publishers, Norwood, MA, 1998.
pp. 239–260

http://www.christopher-gordon.com/

Chapter II

Assessment in Literature-Based Reading Programs: Have We Kept Our Promises?

Tanja Bisesi, Devon Brenner, Mary McVee, P. David Pearson, and *Loukia K. Sarroub*[1]
Michigan State University

We have made incredible progress, both conceptually and practically, in the development of literacy assessment tools that appropriately reflect the goals and activities of literature-based reading programs. This progress, however, has not come without obstacles, many of which have not yet been (and may never be) fully negotiated. The purpose of this chapter is to provide an overview of the "promises" we as a literacy assessment community have made to ourselves, as we implement new forms of assessment for new purposes, and to critically evaluate our progress toward keeping those promises. We begin by briefly describing recent shifts in literacy instruction that have prompted us to make a set of promises for better literacy assessment. Second, we lay out the implicit promises we have made to ourselves as we have worked to develop alternative assessment tools and procedures and judge how well these promises have been kept. Finally, we address dilemmas that we will continue to face as we develop new literacy assessment tools and implement them for new purposes.

AN HISTORICAL LOOK AT LITERACY INSTRUCTION AND ASSESSMENT

To fully appreciate and effectively evaluate progress made and challenges still faced by those who engage in the development of literacy assessments, it is important to situate our discussion in an historical context. Thus, in this section, we begin with a description of shifts in literacy instruction, including a portrait of a classroom in which these shifts have been implemented. We also discuss the challenges to traditional assessment practices presented by these shifts in literacy instruction, challenges that have prompted us to promise ourselves better literacy assessment and to develop an array of alternative assessment tools.

Shifts in Literacy Instruction

The views that we hold concerning what it means to be literate shape the way we teach and assess our students' literacy. Within the last 50 years there has been a shift in the beliefs of the literacy education community, not only in terms of what it means to be a literate individual, but also in how students learn and should best be taught (Langer, 1991). These changes in basic assumptions about literacy, knowledge, teaching, and learning are represented in changes in literacy education, including the recent literature-based instruction movement. At least four specific shifts accompanying this movement have had an influence on the way we think about assessing literacy growth and achievement and are reflected in the following instructional portrait: (a) emphasis on personal response, (b) appreciation of social and cultural aspects of literacy, (c) recognition of the fundamental and interrelated nature of language, and (d) concern for the meaningfulness of materials and tasks.

A Portrait of Current Practice. June is an exemplary, fifth-grade, literature-based reading teacher. Her classroom represents the best of what literature-based reading instruction has to offer in terms of instruction, curriculum, and assessment. In particular, when you walk into her classroom you see students engaging in the kinds of complex, meaningful activities that you would hope to see when visiting a literature-based classroom. Students are reading good literature (e.g., *Hatchet* by Paulsen, 1987; *Number the Stars* by Lowry, 1989) silently and in groups, responding to the literature in written journals, and using these journal responses as a tool for talking about the literature in student-led, small-group and teacher-guided, whole-class discussions. June does not limit her instruction to literacy conventions (e.g., how to decode print; how to talk about books). She also creates opportunities for students to develop comprehension strategies (e.g., summarizing, sequencing, predicting), and provides instruction on literary elements (e.g., genre, text structure), as well as models for responding personally (e.g., "how does this relate to my own life?"), creatively (e.g., "I would change the end of the story to this . . ."), and critically (e.g., "I wonder why the author wrote the book this way?") so that students develop rich interpretations of the literature they read. Further, June uses a portfolio assessment system, including student-generated artifacts (i.e., journal entries, discussion tape/transcripts, journal and discussion self-assessments) and tools that she has created (e.g., anecdotal notes, discussion and journal checklists) to report student strengths and weaknesses and growth to parents, to guide and help students reflect on their own learning, to aid her in instructional planning, and to demonstrate curriculum effectiveness to her administrator. June's classroom, while exemplary rather than typical, illustrates the shifts in literacy instruction we have experienced as a field.

Emphasis on Personal Response. The constructive and personal nature of literacy has been emphasized in literature-based instruction and is instantiated in the tasks (e.g., journal writing) and goals (e.g., personal response) of June's daily instruction. Comprehension instruction of the 1980s concerned itself with the constructive processes involved in textual understanding, and remains a focus (e.g., comprehension strategy goals) of literature-based teachers like June. Its goal, on the one hand, was to help students effectively "get," through activation or building of background, what was "in" the text, the author's message. Literature-based instruction, on the other hand, draws on reader response theories of textual understanding (Langer, 1990; Rosenblatt, 1991) that stress the transactional nature of reading, including what readers bring to the text as well as what they take from it. Instruction based on reader response theory also highlights the evolutionary and personal nature of interpretation, as the reader takes different stances on the text and responds both aesthetically and efferently. This orientation is reflected in June's use of journal writing and discussion as windows on student personal response as well as her direct instruction on a variety of ways to personally respond to text (e.g., "What do I like about the story and why?" "How does this relate to my own life?").

Appreciation of Social and Cultural Aspects of Literacy. With an increasing awareness of the social and cultural diversity of the student population has come a greater appreciation of the sociocultural aspects of literacy. Literature-based curriculum developers (e.g., Raphael, Pardo, Highfield, & McMahon, 1997) have begun to draw on the ideas of Vygotsky (1978) in an attempt to celebrate and cultivate these aspects of literacy. Literature-based programs grounded in a sociocultural perspective assume that knowledge is socially constructed within the context of collaborative, meaningful activities (Brock & Gavelek, Chapter 4). While the comprehension instruction of the recent past (e.g., Pearson & Fielding, 1991) attempted to help students get the author's message and make accurate inferences through the use of specific strategies such as KWL (Ogle, 1986), socially-oriented literature-based instruction provides students with multiple opportunities to demonstrate, internalize, and transform their knowledge and understandings through social interaction with their teacher and peers. These kinds of opportunities for socially-based knowledge construction are clearly demonstrated in our description of June's classroom. For example, June relies heavily on small-group and whole-class discussions as contexts for students to construct diverse and multiple interpretations of text.

Recognition of the Fundamental and Interrelated Nature of Language. Recently developed literature-based programs (e.g., Raphael et al., 1997; Au, Carroll, & Scheu, 1997) have almost universally recognized

the fundamental and interrelated nature of language. While comprehension instruction of the past decade focused primarily on reading and writing and the ways in which each could be used to support the other, literature-based instruction targets oral language processes (e.g., speaking, listening) as a means of constructing text-based understanding and interpretations, as well as supporting written literacy processes. Thus, from the perspective of literature-based instruction, oral and written language promote the development of each other as they both contribute to new forms of thought and learning (Wells & Chang-Wells, 1992). Again, this perspective is reflected in June's instruction, as she has students read texts, write about their thoughts, and share ideas in the context of oral discussion.

Concern for Meaningfulness of Materials and Tasks. The use of "real" literature within the context of meaningful literacy activity has become a staple of literature-based literacy instruction. While comprehension instruction of the past decade stressed the constructive nature of reading, as does literature-based instruction, it relied almost exclusively on contrived textual materials or excerpts from complete literary works to constraint the students' reading task and help them develop effective comprehension strategies. Building on comprehension instruction's constructive nature, which drew on cognitive approaches to literacy (e.g., schema theory) in order to equip students with strategies (e.g., drawing on background knowledge) for understanding texts more effectively, literature-based instruction also promotes having students engage in holistic (e.g., reading a whole novel), meaningful (e.g., talking to peers about books read) literacy activity. These activities, where the teacher (or a peer) provides an appropriate level of assistance or scaffolding, help the student to make sense of text in complex and diverse ways (e.g., from multiple perspectives) rather than relying on a predictable or contrived text structure to provide the support. June's instruction, described above, reflects this concern for meaningfulness, as she has students read complete pieces of interesting literature and supports (along with peers) their evolving understandings.

Challenges to Traditional Assessment Practices

While beliefs about literacy learning and instruction have changed in the ways previously described, the principles and methods that shape the way we assess students' literacy have remained relatively unchanged. Many of the challenges to traditional assessments such as standardized tests have arisen from the fact that these assessments do not align well with literature-based literacy curricula (see Bisesi & Raphael, in press).

Standardized tests tend to tap isolated, low-level skills in decontextualized contexts (Haladyna, Nolan, & Haas, 1991; Shepard, 1989), rather than the complex understandings, personal responses, or intertextual interpretations constructed during meaningful literacy activities. They do not reflect the tasks or texts, such as trade books, used within the context of literature-based instruction. These tests have led students to focus on performance over learning, and they have caused teachers to focus on tested skills rather than their beliefs about literacy (Paris, Calfee, Filby, Hiebert, Pearson, Valencia, & Wolf, 1992; Shepard, 1989). The narrowing of curricula, to focus on the isolated and low-level skills caused by high-stakes standardized testing, is not consistent with the interactive, social, holistic aspects of literature-based instruction. What is uncertain is whether *any* form of multiple-choice, standardized assessment can ever support the goals and principles underlying literature-based reading. Indeed, the most serious attempts to build multiple-choice formats to meet such goals, the statewide assessments in Michigan and Illinois (Valencia, Wixson, Peters, & Pearson, 1989), have met with incredible resistance and criticism from literature-based reading advocates.

In addition to serious problems with curriculum misalignment, conventional tests have been taken to task for a variety of other sins of omission or commission. Tests don't respect the cultural nature of language and literacy (García & Pearson, 1994). They have a long history of negative impact on minorities, with some forms of bias being more blatant, and some more insidious (García & Pearson, 1994). With an emphasis on secrecy and the isolation of individuals during testing, tests have ignored the social aspects of literacy learning (Pearson, DeStefano, & García, in press). With their focus on objectivity and machine-scoring, tests remove the very individuals most responsible for making decisions, the teachers, from the evaluation process.

As we became aware of these challenges to standardized tests (e.g., Paris, Lawton, Turner, & Roth, 1991; Shepard, 1989), we developed an awareness of the need for appropriate, alternative approaches for thinking about, examining, and evaluating students' learning (e.g., Paris, Calfee, Filby, Hiebert, Pearson, Valencia, & Wolf, 1992; Au et al., 1990; Valencia, 1990). This concern prompted us, as a profession, to make a set of promises to ourselves, our students, and our constituents as we went about the business of developing assessments more appropriate to our curriculum and our information needs in the 1990s. In the remainder of this chapter, we discuss these promises and evaluate how well we have kept them.

PROMISES

The set of promises involved authenticity, instructional validity, openness, diversity, client-centeredness, and imaginative assessments.

Authenticity

First was the promise of authenticity. We promised to build assessments grounded in real-world literacy activities and tasks, rather than invented out of psychometric convenience. Our new assessments would have the look and feel of the real thing. Students would read real texts, like the trade books used in June's classroom, not short snippets contrived by item writers to provide lots of tricky distracters or opaque items focusing on obscure details of the text. Instead of selecting the line or words in a snippet that illustrate correct grammar, diction, vocabulary, punctuation, or spelling to demonstrate their knowledge of composition, students would demonstrate both their writing prowess and control over conventions by writing real essays, stories, and reflections, with plenty of time to complete the steps in the writing process—activities emphasized in literature-based classrooms like June's.

Authenticity is complicated. The most important question is the source of the standard for authenticity. If authenticity is determined by comparing a task to a curriculum, then given a separate skills, decontextualized curriculum, many of the decontextualized, specific-skill, multiple-choice items of standardized tests might well meet the authenticity standard. However, if authenticity comes from the uses of literacy in everyday life, or from the concepts of literacy that underlie the curriculum in process writing and literature-based reading, then only performance tasks and work samples in portfolios are likely to meet the authenticity standard. Authenticity is *not* identical to curricular validity. In fact, for an assessment task to meet the authenticity criterion, it must come from a curriculum that is itself grounded in the authentic uses of literacy outside of school settings. This authenticity was illustrated in the portrait of June's instruction, as she encouraged students to discuss the books they read and use their writing as a tool to support their discussions.

How well have we kept the promise of authenticity? Tolerably well, we think. The proliferation of articles and books about portfolio projects in schools throughout the English-speaking world is quite amazing. Our own work in schools in Michigan suggests that an increasing number of schools are moving rapidly to a central role for portfolios, at least for classroom use. Increasingly, multi-draft papers, response journals, and essays reflecting on one's progress in reading and writing are being included in portfolios. And at least a few states, such as Vermont, Kentucky, and Maryland, have developed elaborate portfolio or performance assessment systems for statewide assessment and school accountability schemes.

On the negative side, portfolios and performance assessments are often seen as auxiliary to the "real" assessment system of standardized tests and grading practices. Rarely are they used as a part of the grading process in schools. In fact, some educators disparage the use of portfolios for grading

on the grounds that it somehow makes the experience less valuable, in terms of ownership and reflection, for students. Subjecting student work and reflection to scoring rubrics or point values seems to "sully" the whole process (e.g., Tierney et al., 1991). We find ourselves in a dilemma. We are worried about the negative impact of connecting high stakes (grades and scores) to portfolios. Yet by failing to do so, we may insure a marginalized role for portfolios in the overall assessment picture. At the large-scale level, we note that many states (e.g., California, Wisconsin, and Indiana) have either reversed or abandoned performance assessment systems in response to concerns about expense, time, and intrusiveness.

Instructional Validity

Second was the promise of *instructional validity.* To say that an assessment possesses instructional validity is to say that it both reflects and promotes good instructional practices. The first requirement of instructional validity is to build assessments derived from best practice—not just any old instructional practice, but best practice—or what we as a profession point to when we say, this is exemplary (an example of this kind of exemplary practice is reflected in our portrait of June described earlier in this chapter). A second requirement of instructional validity is that the assessments lead to good instructional practice; teachers who use the assessments will end up exposing students to first-rate instruction. A third requirement of instructional validity that follows from the previous requirements is that assessments will no longer put teachers at odds with their better judgment. As tests have assumed increasingly high stakes for individual students, teachers, or schools, teachers have experienced internal conflict. Many feel compelled to put their instructional programs on hold in order to get students ready to take tests so foreign to their normal curriculum that special preparation is necessary.

To say that we have promised to worry about instructional validity is to say that we have promised to evaluate the validity of tests in terms of their consequences, their *consequential validity,* for students, teachers, and schools (Messick, 1989; Linn, Baker, & Dunbar, 1993). Literacy assessments should result in decisions that propel students into activities that are optimally suited to helping them become more accomplished readers and writers, such as June's literacy portfolios, which encourage students to be self-reflective readers and writers. All too often, we fear, as a consequence of participating in examinations, students are guided into inappropriate or counterproductive activities, such as practicing the skill of filling out bubble sheets. According to this criterion of consequences, such assessments would be judged invalid.

How have we fared on this promise? Again, moderately well. With the rapid spread of portfolios and performance assessments, accompanied by

their use by teachers in making curricular decisions for classes and instructional decisions for individual students, more students than ever before are gaining access to challenging, engaging, curricular tasks, like those that are part of June's curriculum. These tasks involve lots of reading, writing, drama, and discussions prompted by the need to find a diverse array of artifacts to represent student growth and accomplishment. Unfortunately, these positive developments are only part of the picture. We still hear all too many accounts of schools and teachers who feel compelled to spend hours and hours preparing students to take high-stakes performance assessments that will result in published scores for the schools and consequences (such as diploma endorsement) for students. We even hear stories of classrooms in which students and teachers find themselves "packing to the portfolio" (i.e., creating items specifically to enhance the appearance of the portfolio) rather than "teaching to the test." If and where this occurs, it is strong evidence of the insidious power that we, as a society and a school culture, accord to assessments of any and all stripes when we require them to serve high-stakes functions in our society.

Openness

Promising "openness" in our assessment systems means that teachers, students, and parents are privy to the principles and standards underlying the assessment systems that affect students' lives and academic well-being. Assessment can be viewed as a process of making claims about knowledge (Wiggins, 1993). Students, teachers, and parents should have the opportunity to understand (and even influence) the implicit and explicit criteria underlying knowledge claims made in the classroom and in the larger school setting. Consequently, openness incorporates the prior knowledge of all experiences and tasks that teachers and students have in their collective possession (Wiggins, 1993). In addition, openness requires clarity in the explication of standards and criteria that teachers use to make judgments about students' work. In turn, the need for clarity suggests that dialogue among students and teachers about standards should become a central part of the assessment process. Such as is the case in June's classroom when she negotiates with students target performance criteria for journal writing and discussions. In other words, tests and test experiences can no longer be the "secret" journeys that students dread to make because of intimidation and fear of the unexpected. Instead, dialogue and common exploration can render the whole process, including the knowledge requirements and the standards for mastery, transparent. Students and teachers, by having the opportunity to be open about evaluation and assessment procedures and criteria, can easily draw connections between curriculum and assessments derived from it. Openness, then, allows for individual as well as collective opportunities to claim a genuine

understanding both of knowledge and the ways that students can claim that knowledge for themselves.

Openness also means being open to self-examination. Related to this is the issue of accessibility. If the assessment system is not open to both self-examination and public scrutiny, people who can negotiate the system (figure out the rules on their own) are automatically privileged. Those who cannot negotiate the system tend to use compensatory strategies with short-term positive effects and find themselves falling behind; they are then tracked differently from those who can effectively negotiate the system. It is also worthwhile to note another side of openness. Lack of secrecy also makes teachers and students vulnerable to public scrutiny. Standardized tests, for example, expedite the process of large-scale evaluation because they are rather uncomplicated and can "travel" across a range of teaching philosophies and styles (which may mean simply that they are equally as irrelevant to the goals of many philosophies and styles). Performance-based assessments, on the other hand, are often co-constructed by teachers and students (and even parents) and may raise the specter of subjectivity (a positivist assumption that has been imported into the constructivist paradigm). The complicated implicit and explicit agreements embedded in performance assessment can be difficult for teachers or the public to penetrate. One teacher we worked with told us that she could implement the portfolio, for example, as the only means of assessment if she "put herself in her classroom with her students and did not pay attention to outside pressures" (Sarroub, Pearson, Dykema, Lloyd, 1996; p. 13). In other words, the challenge we face in being open is to balance our need to create accountable systems of evaluation that are clear to the public with our desire to acknowledge and respect the perspectives and decisions teachers and students make.

Progress on this promise is mixed. In our attempts to be more "open" in the examination of student work, we have shifted our focus from standardized, multiple-choice assessments to alternative forms such as performance-based assessments. Because measures are derived from actual performance, performance-based assessments are more "open" (and perhaps valid) because students have more opportunities to reflect on and be engaged in their work. In addition, the widespread use of rubrics, especially rubrics that are published and widely shared in advance of actual assessments, provides everyone with access to the apparent standards for mastery. For example, one teacher we worked with showed us his eighth-grade students' papers in which his students focused on four criteria in their writing: content, voice/creativity, form, and mechanics (Sarroub, Pearson, Dykema, & Lloyd, 1996). By utilizing this rubric, the students were able to openly negotiate and reflect on ways to meet the writing standards in their classroom. However, many problems existed. First, it was not always apparent just how "clearly" the rubrics con-

veyed a portrait of mastery. In other words, a good rubric evolved from a teacher's examination of many aspects of students' work over time, such that both the teacher and the student had a clear understanding of the relative difference in mastery between, for example, a 3.5 and a 4.0. This process, however, was a difficult one to navigate, as teachers and students tried out new types of assignments and changed their curricula.

Second, as suggested by Linn, Baker, and Dunbar, "serious validation of alternative assessments needs to include evidence regarding the intended and unintended consequences, the degree to which performance on specific assessment tasks transfers to other situations, including everyday performance, and the fairness of the assessments" (1991, p. 20). In other words, reconciling internal (classroom-based) and external (outside of the classroom domain in which we are currently working) means of accountability is crucial, especially if assessments are to reflect curricular and societal goals. We must be open to the possibility that it is not obvious whether internal or external means of accountability are more open. In fact, students complain all the time about what it is exactly that the teacher expects of them. Hence, the jury is still out on our promise of openness; we need more reflexive, critical examination of current tools before we can make any definitive pronouncements.

Diversity

Even before literacy and teacher educators began to consider the role of assessment in literature-based programs, researchers and those concerned with the education of diverse students had begun to call for assessments that were not biased in favor of mainstream students. Traditional assessments such as standardized tests as well as more specialized measures such as IQ tests, it was argued, were biased toward students with high levels of proficiency in English and knowledge of Anglo mainstream culture. In response, some proposed culture-free or culture-fair testing, while others proposed changes in how tests were norm-referenced, and yet others advocated testing in the student's first language.

As these and other suggestions surfaced, various drawbacks also became apparent. For example, culture-free and culture-fair tests did not eliminate the correlation between IQ performance and socioeconomic status and also had lower predictive validity than traditional standardized tests; changing the way tests were norm-referenced was found to be expensive and problematic because tests could not be compared to the general population; and testing in students' first languages was also found to be problematic because it is difficult to determine which language is dominant for the student and it is extremely difficult for a test to reflect the multiple ways language is used

(García & Pearson, 1994). Yet in spite of these and other difficulties that have arisen, we promised ourselves that our new assessments would value and more accurately and fairly assess the knowledge of students from diverse economic, racial, and linguistic backgrounds. Toward this end, authentic, contextualized assessments have been proposed as alternatives to standardized, formalized measures, and progress has been made in designing various tools such as portfolios, rubrics, standards, anecdotal records, essay exams, and the like. Thus, there are more options available to educators. Viewed from this perspective, we may even conclude that we have fulfilled our promise.

We must be careful, however, not to assume that the availability of such alternatives is a solution in itself, for little research exists to demonstrate that such measures are truly more reflective of what students with diverse linguistic and cultural backgrounds know and can do (Madaus, 1994). At a national level, research that has been done around various performance-based tests, such as the National Assessment of Educational Progress (NAEP), still indicates significant achievement differences between ethnic and linguistic minorities and their Anglo counterparts (Linn, Baker, Dunbar, 1991; NAEP, 1994). At the local level, there is little documentation that explores the implementation and sustained use of alternative assessments with students from diverse cultural, economic, and linguistic backgrounds, although there are a few notable exceptions.

Central Park East Secondary School and International High School are among those schools that have successfully restructured to include portfolio assessment as an integral and successful component in the curriculum (Darling-Hammond, Ancess, & Falk, 1995). These successful examples indicate that performance-based assessment can provide an equitable accounting of what students from diverse cultural and linguistic backgrounds have learned. We must keep in mind, however, that performance assessment is not the only issue that precipitated success in these school settings, as many other factors also came into play as these schools were restructured. While results from particular schools are encouraging and provide a model for future reform, the promise that alternative assessments are more equitable must continue to be viewed from a critical stance at both the local and national level.

From a practical standpoint, and as exemplified by Central Park East Secondary School and International High School, the use of innovative assessments requires innovative thinking by teachers, administrators, and policy makers. To fulfill the promise, teachers, administrators, and policy makers must have the knowledge and resources to implement alternative assessments. In part, this means that teacher educators and researchers interested in assessment must provide resources, support, feedback, and documentation for those who are willing to implement alternative assessments. It also means

that we will need to commit ourselves to valuing alternative assessments so that high-stakes decisions within schools reflect these changes.

We must acknowledge that implementing and sustaining alternative assessments with learners from diverse backgrounds will not help those learners unless we also work to develop a reflective stance among key participants—parents, students, teachers, administrators, and policy makers. This is of particular importance for teachers because, while the student population has continued to increase in linguistic and ethnic diversity, the teaching force has remained predominantly Anglo, female, and middle class. Thus, it is not only necessary to develop and implement alternative assessments, but to facilitate teachers' understanding and reflection by considering questions such as the following: What are the benefits and pitfalls of alternative assessments? Are alternative assessments less biased toward linguistic and ethnic minorities? Do they allow students to draw on their cultural and linguistic strengths? What are the practical concerns teachers must consider when implementing and sustaining such alternatives? Only after we have fully explored these and other related questions will we be able to fully evaluate our progress toward creating assessments that value ethnic and linguistic diversity.

Client-Centeredness

As our beliefs about the nature of literacy learning and instruction and the nature of assessment changed, so did our commitment to the clients of assessment. We vowed that our new assessments would be more client-centered and provide valuable information to all audiences: students, teachers, families, administrators, and policy makers. In the past, the assessments that mattered were designed for and focused on the needs of administrators and program decision makers (Abruscato, 1993). Alternative assessments, we vowed, would provide new and more important information to teachers, students, and parents. In this section, we talk about the various clients for assessment and the ways that alternative assessments have attempted to meet their needs.

Teachers need information that helps them describe what their students can do, that allows them to discover and recreate the kinds of experiences that help children learn, that encourages them to modify and refine their teaching practices, and that facilitates their communication with students about both expectations and accomplishments. Traditional high-stakes assessments, such as report cards and test scores, often do not provide such information. On-going classroom activities have the potential to provide information about progress on important curricular goals, especially when they are supported by the ancillary tools of alternative assessment: reflection, collaboration, rubrics, and portfolio collections.

Students need assessments that allow them to reflect upon their learning, to gauge their progress, and to set future goals for learning (IRA/NCTE standards, 1994). All of this can take place when the act of performing the assessment itself causes students to think actively and reflect on their own learning. In June's classroom, for example, students reflect on their growth and accomplishment as they collect, write about, and share their portfolio artifacts.

Teachers can also help students to understand their own learning progress and to set learning goals by conducting assessments that serve students' needs as well as their own. As teachers collect and think about the artifacts they need to understand their students and their own teaching, they can share what they learn with students. Rather than telling students "how they did" on a task, teachers can share their assessment tools—such as benchmarks and rubrics—and their interpretations with students in order to assist students in evaluating their own efforts and making decisions about their future learning. In June's classroom, for example, June regularly scores her students' portfolios according to a set of target benchmarks, and shares those benchmarks and scores with students, talking with them at length about just how and why she assigned the score she did.

Parents are often most interested in information that lets them understand their own child's progress. Traditionally, parents have been provided with test scores, letter grades, and unannotated work samples. Alternative assessments often provide parents with unfamiliar kinds of data, including student and teacher reflections on growth and accomplishment and information about the process, as well as the products, of assessment. While teachers are working to provide parents with richer, more detailed descriptions of student learning, parents often continue to ask for data that allows them to make comparisons between students. When faced with such questions, teachers reluctant to make comparisons can help parents understand just how their child "stacks up" against predetermined and preset benchmarks, rather than against other students. Teachers can share with parents their standards for learning. They can help parents have a sense of the typical or a sense of the ways that standards and expectations grow and change across time and across grade levels.

Have we met these promises? It isn't clear at this point that assessments are being used consistently in ways that meet the needs of students, teachers, and parents. Gillespie, Ford, Gillespie and Leavall (1996), for example, raise concerns about portfolio assessment, including concerns that teachers may tend to focus on management and collection rather than reflection and learning, and that portfolios may lead to less, rather than more, conversation with teachers.

However, new kinds of assessments and new attitudes toward assessments have the potential for serving these three types of clients (Bisesi, 1997).

Teachers have begun to add their voices to the assessment milieu and have become actively involved in developing and creating alternative assessments for use in their own classrooms (Wiggins, 1993). The documents that arise naturally out of sound instructional practices (e.g., written journals, responses to literature, performances) have taken on new significance as artifacts for assessment. Increasingly, teachers, students, and parents pore over these artifacts to create portfolios that reflect the learning that has taken place (e.g., Graves & Sunstein, 1992; Educators in Connecticut's Pomperaug Regional School District 15, 1996). The process of creating these collections has been as valuable as the information that arises out of them.

Parents, teachers, and students are grouped together here in part because the emphasis on these assessment audiences has increased, but also because all of these clients, in their own ways, needs thick description of students' learning and accomplishment. Administrators and policy makers need, or at least request, aggregated data, which often takes the form of scores averaged across students, classrooms, or schools. But even though administrators continue to feel the need for numbers and scores, the assessments that lead up to those scores are changing. Teachers and administrators in several states (e.g., Vermont, Kentucky, California, Maryland) have worked together to provide both individual and aggregate scores that arise from portfolios or performance assessments. These assessments have the potential of providing results that support instruction and also permit policy analysis, thus simultaneously supplying valuable information not only to administrators but to teachers, students, and parents. It is possible that the same tools that teachers, parents, and students reflect on could be consistently and concisely used to provide information to audiences outside the classroom. While there are inherent difficulties in exporting classroom-based assessments (Pearson, DeStefano, & García, in press), both in maintaining the richness of information and in preserving worthwhile instructional activities, this process holds great potential for all the clients of assessment.

Imaginative, "Break-the-Mold" Assessments

Standardized, limited-response tests, including large-scale norm-referenced as well as curriculum-oriented criterion-referenced tests, have been the touchstone of educational assessment for several decades. While these tests have come to serve the purposes of school administrators and policy makers quite well, providing a simple measure of achievement across large numbers of students and a degree of fairness with their standardized "level playing field" administration and scoring, they have become increasingly limited in helping teachers, students, and even parents understand the complexities of curriculum, instruction, and student learning (Bisesi, 1997; Farr, 1992). New

forms of assessment came to us with the potential not only to provide information that would allow us to monitor student academic progress, but also help teachers to teach better and students to learn more effectively.

Given this potential, we promised ourselves to develop new, imaginative, "break-the-mold" assessments that we did not assume or require to resemble or even correlate highly with conventional assessments. We started to ask ourselves questions such as, "What would literacy assessments look like if we put the emphasis on literacy rather than assessment?" "What would these assessments look like if reading theory rather than psychometric theory were the driving force behind their development?"

The process of literacy teaching and learning, as it is conceptualized both theoretically and in today's literature-based classrooms, is awesome and complex. In literature-based classrooms students engage in a range of purposeful and interrelated language arts activities including reading, writing, and talking about good literature and trade books, as instantiated in June's instruction described earlier in this chapter. Students interpret and reinterpret the texts they read in written forms and through discussion with both teacher and peers. These rich activities, within which students are making sense of text, require rich and imaginative assessment tools, in order to provide information that reflects the range of student learning taking place (e.g., Wiggins, 1993; Resnick & Resnick, 1985).

Our promise for "break-the-mold" assessments, not limited to the standardized test models of the past, has been relatively well kept. The assessment literature is full of accounts of attempts, on both a large-scale (e.g., NAEP; Vermont; Kentucky) and classroom-oriented level (Bisesi, 1995; Graves & Sunstein, 1992; Tierney et al., 1991), to create portfolios and performance assessments that apply new methods of collecting (e.g., Bisesi & Raphael, 1996; Au et al., 1990) and interpreting information about literacy teaching and learning (e.g., Moss, 1996; Delandshere & Petrosky, 1994) and new standards for judging the validity of our methods (e.g., Moss, 1992; Messick, 1989). We must, however, craft our assessments carefully to ensure that the focus remains on curriculum and not on the "test." While some researchers have advocated driving curriculum/instruction in directions we want it to go by creating and implementing "good" assessments (Resnick & Resnick, 1985), we have to take care not to design innovative assessments that lead curriculum too far off track. Thus, we should look to curriculum for our lead in developing break-the-mold assessments.

ONGOING DILEMMAS

We have talked about the ways that the field of assessment has responded to changes in literacy instruction and about criticism of more traditional assess-

ments. Not all of the promised innovations have materialized yet. But the promises of authenticity, validity, respect for diversity, imagination, and client-centeredness continue to be ideals that parents, teachers, administrators, and students strive for as they continue to reflect on and articulate their assessment needs. Even if we could fulfill our promises by expanding the tools and the range of evidence for assessing students, we still would not reach a fully satisfying outcome, for a set of dilemmas would remain. These dilemmas are not unique to alternative assessments; rather they are inherent in any act of assessment. We highlight four that are as important as they are elusive: (1) feasibility, (2) fairness, (3) purpose, and (4) values.

Feasibility

Feasibility of assessments concern their benefit-to-cost (in terms of time and money) ratio. While traditional, standardized tests are cost effective, they have imposed on the instructional and learning time of teachers and students, without benefiting the teaching and learning taking place in literature-based classrooms. New assessments promise to provide more meaningful information about teaching and learning, while intruding less on (or even enhancing) the everyday literacy activity taking place in the classroom. Because portfolios draw on documents generated during the course of instruction, assessment information is directly relevant to decisions about teaching and learning. Furthermore, assessments are more feasible for teachers to implement in their classrooms and school buildings (e.g., Au et al., 1990).

Nevertheless, using instructionally related documents can place a burden on teachers and schools, a burden that only time and effort can carry. First, standards for evaluating documents must be clearly defined through extensive negotiation and consensus building to avoid later disagreements and disappointment. These standards are frequently laid out in the form of rubrics to simplify the process of evaluating student work (e.g., Bisesi, 1996). Second, because interjudge reliability is often low for artifacts scored using rubrics, the number of artifacts in a portfolio collection will need to be relatively high in order to ensure adequate representation of performance.

Fairness

Fairness, being as equitable as possible to all who must live with tests, has long been touted as an ideal, a standard, of all good assessments. It is fairness that drives us to "treat everyone equally." But equitable treatment is an elusive goal, and when we attempt to ensure it, we often end up victims of a conspiracy of our own good intentions. The problem is that any time we use one criterion (e.g., everyone answers the same questions under the same con-

ditions) to establish "equity," then some individuals, those who would have benefited from a different criterion of "sameness," are marginalized. Falling victim to our own good intentions is all the more serious and all the more likely with respect to matters of cultural and linguistic diversity.

In discussing issues of equity, it is common for us to use metaphors of equality, such as a level playing field or a common yardstick. Yet, the one-size-fits-all approach is likely to perpetuate the differences in academic performance that we commonly find in indices such as dropout rates, scores on college entrance tests, and national standards attainment. Put differently, the level playing field approach establishes one kind of equity (they all did the same task under the same conditions) while allowing other kinds of equity (e.g., the opportunity to perform familiar tasks in familiar contexts or the opportunity to put one's best foot forward) to vary dramatically. Ironically, in the few documented instances in which we have used assessment tools that recognize, acknowledge, and value diversity, we get a very different and more positive picture of students' capabilities (see the earlier section on diversity).

If we want to establish an alternative type of equity, an equity in which all students get the opportunity to put their "best-foot-forward" or "show their stuff," then other options may be necessary. The best-foot-forward metaphor for equity would lead us toward *choice* as a primary tool for achieving equity: choice of passages to read, questions to answer, prompts to write to, projects to complete, or even sociolinguistic contexts in which to work. In principle, this would not seem to be a problem within a performance assessment milieu; performance assessment, particularly portfolios, ought to allow—even champion—diverse ways of solving problems, accomplishing tasks, and meeting standards.

Even staying within the logic of the level-playing-field metaphor, other options are available. If we use the framework of dynamic assessment, we change the task for ourselves as teachers and our students. We end up asking ourselves how much support is needed to help particular students accomplish a specified goal or level of achievement. In this instance, instead of leveling opportunity, we are leveling achievement and allowing the type and amount of scaffolding provided to vary. Consider the revolution that might occur if choice and scaffolding rather than standardization drove our quest for equity. We would have a very different concept of assessment, not to mention a very different concept of curriculum.

Purpose

As Farr (1992) so astutely pointed out, different assessment audiences need different kinds of assessment information. Unfortunately, there is often a

tension among the needs of different assessment constituencies. While administrators and other policy makers tend to need aggregated, standardized test score data in order to make decisions about curricular programs or educational policy, and teachers and students need instructionally relevant information to guide curriculum, instruction, and learning in the classroom, parents want information concerning how their individual child is performing relative to classroom standards and his or her peer group (Bisesi, 1997). Sometimes the needs of more powerful constituencies (e.g., administrators' and policy makers' need for standardized test information) subvert the needs of less powerful constituencies (e.g., the need of students and teachers to be held accountable for instruction and learning relevant to literature-based reading programs).

Unless we come to terms with the fact that different assessment audiences need and deserve different kinds of assessment information, the dilemma of conflicting assessment purposes will continue to plague the implementation of alternative assessments. We must respect the unique needs of all assessment audiences and work to build assessment systems that ensure that all groups concerned get the assessment information they want and need (Bisesi, 1997). Furthermore, it is critical that we not fall into the trap of relying on only a single source of information (e.g., standardized tests) to make important decisions about curriculum and instruction (in place in literature-based reading classrooms) in order to avoid undervaluing the complexity of learning we want to encourage.

Values

Literature-based programs are meant to do much more than help students learn to decode text. As they interact with texts and each other, children are meant to become critical, problem-solving users of language. As the authors of the *International Reading Association/National Council of Teachers of English Standards for the English Language Arts* (1994) state, "Our aim is to ensure that all students develop the literacy skills they need to succeed in school, in the workplace, and in the various domains of life" (p. B). All of the other aims of literacy education, from personal response to literature to multiple ways of representing and sharing interpretations of text, eventually link back to helping students grow the literacy skills they need for success in their lives. And because of this link, our final dilemma will never disappear.

While we all hope for successful lives for our students as they use their literacy skills, we cannot ever agree about just what it means to be successful. Defining success, like defining any other benchmark of accomplishment, is an inherently value-laden action. Success might be financial, it might be emotional, it might be intellectual. This discussion of the value-ladenness dilemma provides a convenient and appropriate place to end our discussion

of promises and problems. Differences in values color every aspect of the assessment of student learning in literature-based classrooms. The setting of goals, the choosing of evidence, the interpretation of artifacts, all ultimately depend upon teachers' values and their goals for teaching and learning. It is perhaps because assessment is so value-laden that the promises we have made—promises for authenticity, instructional validity, diversity, openness, client-centeredness, and imagination—have been so difficult, and so important, to keep.

CONCLUDING COMMENT

In conclusion, we as a literacy assessment community have come a long way toward keeping our promises to create alternative assessments that are authentic, instructionally valid, innovative, open, client-centered and respectful of the ethnic, cultural, and linguistic diversity of students making up current literature-based reading classrooms. And while we have kept some of the promises to a greater degree (e.g., the promise for break-the-mold assessments) than others (e.g., the promise for assessments that respect diversity), we must remain committed to these promises, as well as to the ongoing dilemmas of feasibility, fairness, purpose, and values, as we continue to design and implement alternative assessments for use in literature-based reading classrooms.

ENDNOTE

1. Authorship order was determined by attributing first author status to the individual that we, collectively, decided had made the greatest contribution to the overall effort. The second through the fifth authors, we determined, contributed equally to the process, so we listed their names alphabetically.

REFERENCES

Abruscato, J. (1993). Early results and tentative implications from the Vermont Portfolio Project. *Phi Delta Kappan,* 474–477.

Afflerbach, P. P., & Johnston, P. H. (1993). Writing language arts report cards: Eleven teachers' conflicts of knowing and communicating. *The Elementary School Journal, 94*(1), 72–74.

Au, K. H., Scheu, J. A., Kawakami, A. J., & Herman, P. A. (1990). Assessment and accountability in a whole literacy curriculum. *The Reading Teacher, 43,* 574–578.

Bisesi, T. (1997). *Exploring the value of a performance-based assessment: A social validity analysis.* East Lansing, MI: Unpublished dissertation, Michigan State University.

Bisesi, T. L. (1996). Upper-elementary students' written responses to text: A holistic scoring rubric for evaluating journal entries. In D. J. Leu, C. K. Kinzer, & K. A. Hinchman (Eds.), *Literacies for the 21st century: Research and practice* (pp. 76–87). Chicago, IL: National Reading Conference.

Bisesi, T. L., & Raphael, T. E. (1997). Assessment research in the Book Club program. In S. McMahon & T. Raphael (Eds.), *The Book Club project: Exploring literature-based reading instruction* (pp. 184–204). New York: Teachers College Press.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work.* New York: Teachers College Press.

Delandshere, G., & Petrosky, A. (1994). Capturing teachers' knowledge: Performance assessment and post-structuralism. *Educational Researcher, 23*(5), 11–18.

Educators in Connecticut's Pomperaug Regional School District 15 (1996). *A teacher's guide to performance based learning and assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Farr, R. (1992). Putting it all together: Solving the reading assessment puzzle. *The Reading Teacher, 46*(1), 26–37.

García, G. E., & Pearson, P. D. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 20) (pp. 337–391). Washington, DC: American Educational Research Association.

Gillespie, C., Ford, K., Gillespie, R., & Leavall, A. (1991). Portfolio assessment: Some questions, some answers, some recommendations. *Journal of Adolescent and Adult Literacy, 39*(6), 480–491.

Graves, D., & Sunstein, B. (1992). *Portfolio portraits.* Portsmouth, NH: Heinemann.

Haladyna, R., Nolan, S. B., & Haas, N. (1991). Raising standardized achievement tests scores and the origins of test score pollution. *Educational Researcher, 20*(5), 2–7.

International Reading Association/National Council of Teachers of English (1994). *Professional summary: Standards for the English language arts.* Urbana, IL: National Council of Teachers of English.

Langer, J. A. (1990). Understanding literature. *Language Arts, 67,* 812–816.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20,* 15–21.

Madaus, G. F. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's testing policy. *Harvard Educational Review, 64*(1), 76–95.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement in Education.

Moss, P. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher, 25*(1), 20–28, 43.

Moss, P. (1992). Validity in educational measurement. *Review of Educational Research, 62*(3), 229–258.

National Assessment of Educational Progress (1994). *1994 Reading Report Card for the Nation and the States.* Washington DC: National Center for Educational Statistics.

Ogle, D. M. (1986). K-W-L: A teaching model that develops active reading of expository text. *The Reading Teacher, 39*(6), 564–570.

Paris, S., Lawton, T., Turner, J., & Roth, J. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20*(5), 12–20.

Paris, S., Calfee, R., Filby, N., Hiebert, E., Pearson, P. D., Valencia, S., & Wolf, K. (1992). A framework for authentic literacy assessment. *The Reading Teacher, 46*(2), 88–98.

Pearson, P. D., DeStefano, L., & Garciá, G. E. (in press). Dilemmas in reading assessment. In C. Harrison & T. Salinger (Eds.), *Assessing reading 1: Theory and practice.* London: Routledge.

Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (pp. 819–860). New York: Longman.

Raphael, T. E., Pardo, L. S., Highfield, K., & McMahon, S. I. (1997). *Book Club: A literature-based curriculum.* Newton, MA: Small Planet Communications.

Resnick, D., & Resnick, L. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher,* 5–20.

Rosenblatt, L. M. (1991). Literary theory. In J. Flood, J. M. Jensen, D. Lapp, & J. R. Squire (Eds.), *Handbook of research on teaching the English language arts* (pp. 57–62). New York: Macmillan.

Sarroub, L. K., Pearson, P. D., Dykema, C., & Lloyd, R. (1996, December). When portfolios become part of the grading process: A case study in a junior high setting. Paper presented at the annual meeting of the National Reading Conference, Charleston, SC.

Shepard, L. (1989). Why we need better assessments. *Educational Leadership, 46*(7), 4–9.

Tierney, R. J., Carter, M. A., & Desai, L. E. (1991). *Portfolio assessment in the reading-writing classroom.* Norwood, MA: Christopher-Gordon.

Valencia, S. (1990). A portfolio approach to classroom reading assessment: The whys, whats, and hows. *The Reading Teacher,* 338–340.

Valencia, S., Wixson, K., Peters, C., & Pearson, P. D. (1989). Theory and practice in statewide reading assessment: Closing the gap. *Educational Leadership, 47*(7), 57–63.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

Wells, G., & Chang-Wells, L. (1992). *Constructing meaning together.* Portsmouth, NH: Heinemann Educational Books.

Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* San Francisco: Jossey-Bass Publishers.

CHILDREN'S BOOKS CITED

Lowry, L. (1989). *Number the stars.* South Holland, IL: Yearling Books.

Paulsen, G. (1987). *Hatchet.* New York: Puffin Books.